Capacity Management for Cloud Computing: A System Dynamics Approach

Full Paper

Derek L. Nazareth
University of Wisconsin-Milwaukee
derek@uwm.edu

Jae Choi
Pittsburg State University
jchoi@pittstate.edu

Abstract

As the demand for cloud computing as a preferred computing architecture grows, the need for effective capacity planning by cloud providers becomes crucial for their long term viability. Situations involving under-capacity and over-capacity represent lost opportunities and increased overhead. Economic conditions play a critical role in determining the capacity, cost, and revenue of cloud-based services. Using a system dynamics approach, this study evaluates the different conditions in cloud ecosystem from a capacity planning and management perspective, with a view to providing cloud service providers guidance for cloud capacity management.

Keywords

Cloud computing, capacity planning, simulation, system dynamics

Introduction

In surveys of critical information technology issues for managers, cloud computing has been ranked highly (Luftman and Zadeh 2011). Not surprisingly, published statistics indicate that cloud computing successfully attracts many companies as it enables them to pass the demands of managing hardware and software to the third party companies (Oredo and Njihia 2014). The cloud computing market is growing rapidly, with the market expected to top $160B by 2020, starting from a modest $6B in 2008. The size of the market devoted to small and medium business is expected to be around $35B by 2018, with approximately $12B devoted to Infrastructure as a Service (IaaS) products. These figures do not include clouds designed to support individuals, since much of those services are provisioned by the vendors that directly support those services, e.g. Google, Facebook, Apple, etc. Although cloud computing attracts greater attention in the academic literature as it reshapes the IT industry (Marston et al. 2011), extant literature has mainly focused on challenges related to effective adoption from the perspective of consumers. The pressing issues that cloud consumers face, when contemplating the adoption and selection of appropriate cloud service providers and products, mainly stem from vendor side factors like revenue, cost structure, capacity building, service quality, just to name a few. Given the wide variety of problems that need to be addressed, cloud service providers need to consider a portfolio of strategies to properly respond to the mandated requirements, each with different costs and potential benefits.

Cloud service providers often have to assess the consequences of changing economic conditions and decisions about cloud capacity related investments in an a-priori manner, wondering if the capacity decisions made were the appropriate ones. A modeling-based approach that captures the complexities of cloud ecosystem while permitting investigation of alternative economic conditions would provide greater insight for the managers at cloud service providers. This study adopts a design science research methodology (Hevner et al. 2004), developing a dynamic model that allows cloud service providers to examine the effects of different economic factors on the overall cloud echo system.

The rest of the paper is organized as follows. The theoretical background that serves as the foundation for building a cloud capacity model is presented in the next section. A dynamic model of the mechanics underlying the impact of changing economic environment and the implications is assembled and presented next. Results of running the model under a variety of economic conditions are presented. Research and managerial implications of the simulations round out the paper.

Prior Research

Several service models of cloud computing have been identified, including Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS), and Data storage as a Service (DaaS) (Dillon et al. 2010). SaaS enables applications to be accessed through networks as evidenced on Salesforce.com and Google Docs. PaaS offers application development platforms with which users develop cloud services, e.g. Google App Engine. IaaS and DaaS are hardware-based services like Amazon's EC2 and S3.

Previous studies have investigated different aspects of cloud computing such as trust, security, adoption in the context of advantages and risks, to name a few. A research stream on trust attempted to identify ways to enhance trust development between cloud consumers and cloud service providers (Lynn et al. 2016). The area of innovations in cloud security has also been studied (Khansa and Zobel 2014). Other studies focus more on cloud adoption in the context of cloud-based data backup, cloud ERP, and cloud enterprise systems, identifying convenience (Menard et al. 2014), benefits and challenges (Peng and Gala 2014), risks (Dutta et al. 2013), and advantages (Marston et al. 2011). Despite the importance of the topic, however, there are comparatively few published studies on the relationships among cloud capacity, cost structure, revenue structure, and other factors relevant to cloud service provider's perspective.

Cloud Capacity Planning

Cloud capacity can be viewed from two perspectives – the cloud consumer, and the cloud provider. Given the elastic nature of cloud provisioning, it has been suggested that cloud capacity planning from the perspective of the cloud consumer is irrelevant. On the face of it, there are merits to the argument. If additional resources are needed temporarily, they can be acquired, albeit at a cost. However, this assumes that additional resources are provided at identical rates. Most cloud providers offer tiered services, based on volume of resources consumed. Accordingly, cloud consumers should engage in a capacity planning exercise to select the most appropriate tier. This exercise should consider standard and peak demand, and should factor in anticipated growth in the demand. It should also consider the inclusion or removal of applications into the cloud computing environment.

This research focuses on the capacity planning from the cloud provider's perspective. Two contrasting approaches for data center design and operation characterize the industry. The approach favored by providers of dedicated applications, e.g. Google, Apple, Facebook, etc., involves running datacenters at full capacity around the clock, irrespective of demand levels, in an effort to ensure that capacity will always be available in the event of a demand surge (Glanz 2012). The need for instant availability under any demand surge typically leads to the waste of 90% of the energy consumed by these datacenters.

A second approach is favored by cloud providers who support a host of varied customers during a diverse set of applications (Kirk 2016). In this case, the core assumption is that not all customers will manifest their peak demand at the same time. Therefore, they are able to support a large number of customers with a given set of resources, even though their individual peak demand will far exceed the normal demand. The elastic provisioning of resources can be facilitated by autonomic management of the cloud resources (Menasce & Ngo 2009).

Early models for cloud provider capacity planning tended to focus on cloud storage provisioning (Wu et. al. 2010), since this represented the bulk of resources needed by cloud datacenters. With storage costs dropping, and coupled with the explosion in data stored, this approach has ceased to be effective, and a focus on processing becomes more realistic. This is also confirmed by pricing models from various cloud providers, wherein the computing costs now outstrip the storage costs. A planning model based on the provision of different sized cloud virtual machines is presented in (Mao et. al. 2010), patterned on the Microsoft Azure cloud. They employ simulation using a predefined demand pattern to evaluate the effectiveness of their approach. Capacity planning through the use of demand sensitive pricing is discussed in (Sharma et. al. 2011a, Sharma et. al. 2011b). They propose a model that is based on Amazon's EC2 system, and utilize various pricing options to provide elastic resources in the face of shifting demand.

Capacity planning using a load prediction model is presented in (Saripalli et. al, 2011). A number of different models are discussed and a cubic spline approach is utilized. This approach works reasonably well for curve fitting existing data, but is not useful for dealing with changes to external conditions that may affect demand. Another load-based capacity planning approach is discussed in (Jiang et. al. 2012). In this case, seven different forecasting models are presented and compared using a given demand pattern. A cost function is utilized to evaluate the different models. Four different de-provisioning strategies are also evaluated, with a long-term planning window providing the best results. The notion of capacity planning through cloud monitoring is discussed in (Aceto et. al. 2013). However, no specific models or strategies are provided, and the focus remains on the assessment of a host of cloud monitoring tools.

Need for an Alternative Approach

The bulk of the models for cloud capacity planning tend to focus on the narrow issue of provisioning resources in an elastic manner, given a specific demand profile. As such, they address operational-level decisions, using a specific set of cloud resources. A cloud computing manager, on the other hand, needs to address more macro-level planning of cloud capacity, taking into consideration economic factors, overall market size, growth, and resource management, including cost containment. Cloud pricing is also a factor, since it affects market share.

Unlike other quantitative models of cloud computing, this study takes the position that a cloud ecosystem is a complex and dynamic system encompassing many closely related and circularly coupled constructs. Diverse factors and dynamic relationships among them need to be investigated. The dynamic aspects of the cloud ecosystem can be captured through simulation studies. Several simulations options are available for understanding the dynamic aspects of cloud ecosystem, including discrete event simulation, continuous simulation, system dynamics, and agent-based simulation, among others. System dynamics uses a combination of first order linear and non-linear difference equations to relate qualitative and quantitative factors within and across time periods (Sterman 2000) and is based on principles developed by Forrester to study managerial and dynamic decisions using control principles (Forrester 1961).

This research uses system dynamics to investigate the effect of different decisions and conditions on the cloud ecosystem. System dynamics was chosen for the simulation as it permits examination of relationships between constructs within a time period, as well as across time periods. It describes a model that examines the cloud market environment, processing and storage requirements on servers, a range of relevant costs (e.g. energy costs, amortized costs, lease costs, personnel costs etc.), cloud revenue, cloud price, etc. It provides managers at cloud service providers with the ability to investigate the effect of economic factors and their impact under a variety of different scenarios. While the model cannot cover all scenarios, it does provide managers with insights into the relative tradeoffs under different scenarios. This research adopts a design science methodology using the system dynamics model as the artifact of interest. The utility of the artifact is demonstrated through execution of the model using a set of diverse scenarios, and using the results to guide cloud provider managers about potential capacity planning decisions they face.

Cloud capacity model

The cloud capacity model is driven by environmental factors that shape the size of the cloud market. Several providers compete for this demand, and based on income and substitution effects, a specific demand for an individual provider is derived. The model focuses on the provision of IaaS services only. It is acknowledged that a cloud provider will likely provide web support and business applications, but these are not included in the model. Based on the demand for IaaS services, the resources needed to provide these services can be estimated. Costs and revenue streams associated with this demand can be computed. The model assumes a reactive aspect to cloud pricing – a drop in income leads to a reduction in the price, thereby capturing market share to offset shrinking income. As a result, the model is expected to reflect some corrective behavior, rather than smooth performance.

The model was developed over several rounds of iteration and testing, and is depicted in Figure 1. Variables in rectangles represent stocks that can accumulate or deplete over time. Stocks are affected by flows, which are represented by a double arrow and valve symbol. They address incoming or outgoing rates of change. Other variables on the diagram represent converters, which have values that are specified for the given time period. Values of converters are determined by differential equations and are affected by other converters through connectors. Positive signs on connectors indicate that an increase in one will lead to an increase in another. Negative signs indicate the opposite. Reinforcing (all positive signs), or balancing (at least one negative sign) loops are achieved when a chain of connectors traces a path back to the originated converter. Reinforcing loops can eventually generate zero or infinite values for the involved converters while balancing loops generate oscillatory trends, which leads to possible equilibrium.

The methodology adopted for the assembly of the cloud computing model comprised iterative steps of model construction, model structuring, model calibration, and model validation. Since system dynamics models are not deterministic in nature, an iterative build-calibrate-test strategy was adopted.

Model Construction

It is recommended that a grounded theory approach be adopted for the construction of system dynamics models (Luna-Reyes & Anderson 2003). Accordingly, the literature in cloud computing diffusion, cloud data center location, cloud data center layout, and cloud product pricing, was systematically examined to identify constructs that would shape the model. System dynamics models that addressed cloud computing were also studied for relevant constructs. Linkages between constructs were assembled based on findings from prior research. Given the paucity of prior research in the area, additional constructs were needed to complete the detailed model of cloud provider performance.

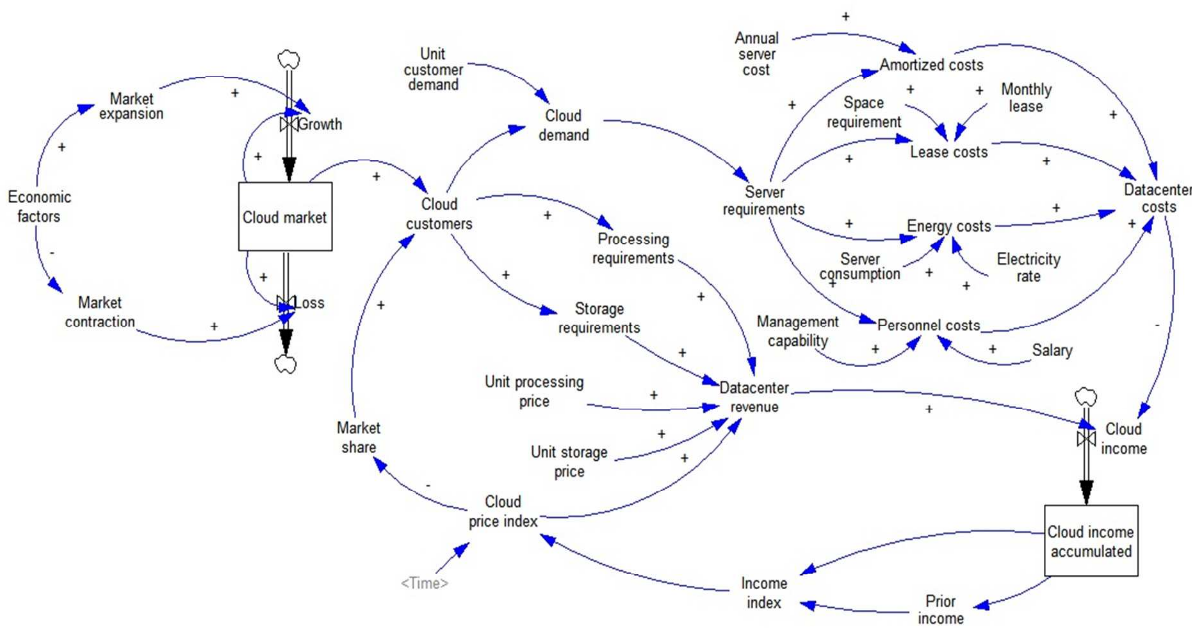

Figure 1. Cloud Capacity Model

Model Structure

The segment on the cloud market conditions is described first. Economic factors shape cloud market expansion and contraction. A growing economy is expected to result in expansion of the overall cloud market as more businesses seek cloud services. A stagnant economy will also result in growth of the overall cloud market, though not as much. This reflects the fact that businesses are looking to outsource their data center operations, and thereby cut costs to stay competitive in the stagnant economy. Only in the case of a shrinking economy will the overall cloud market shrink, as businesses that cannot stay afloat will also obviate their need for cloud services.

The overall cloud market and the market share of an individual provider determine the number of customers the provider can service. The number of customers drives the cost and revenue streams for the provider. On the revenue side, it determines the storage and processing requirements, both of which represent separate revenue streams, and are used to determine the overall revenue generated by the cloud datacenter. The number of customers also serves as the basis for determining the cloud demand, which is met through a number of servers. The servers represent a critical factor in the computation of cloud datacenter costs. Assuming a server life of four years, the amortized cost of the servers can be estimated (Larumbe & Sansò 2013). Given a cabinet configuration for the servers, plus overhead for aisles and other storage, the size of the datacenter can be estimated (Rasmussen 2015). This permits the computation of lease costs, assuming that the facility is leased rather than owned. Another significant cost is the energy costs associated with running the datacenter (Koomey 2015). Using the number of servers and the configuration, estimates for power consumption can be derived, including cooling requirements (Rasmussen 2015). Based on estimates for the number of servers that a data center administrator can manage (Miller 2009), personnel costs for the facility can be computed.

The rest of the model is fairly straightforward. Cloud datacenter income is computed based on revenues and costs. A reactive pricing policy is adopted, wherein the price in the subsequent period is determined based on the current income. The new price is then used to estimate the market share for the next period. This completes a reinforcing feedback loop. If not managed properly, a reinforcing loop in a systems dynamics model can drive the values to zero or infinity. However, rigorously constructed models do not manifest such behavior. Another large feedback loop consisting of cloud demand, server requirements, datacenter costs, cloud income, cloud price index, and market share, functions as a balancing loop, and helps counteract the reinforcing loop. This loop embodies a novel self-correcting mechanism involving price and market share. As market share drops, a price cut is employed to correct the situation. Likewise, increases in market share leave the provider vulnerable to over-capacity conditions when the market inevitably pulls back. To counter this, another price correction is applied. Since the time period chosen was a month, all costs and revenues were estimated on a monthly basis for consistency.

Model Calibration and Verification

Several segments of the model are deterministic in structure and can be calibrated and tested fairly easily. This includes the cost computations, the revenue estimation, and the income assessment. Using published data for the coefficients of various inputs, these segments of the model are easily calibrated and verified. However, it is important to ensure that cost and income ratios are in line with standards in the cloud computing industry. Running the model under a variety of scenarios confirmed that these ratios were consistent with industry trends.

Other segments of the model are a little more complex, and need greater care to calibrate and verify. These include the price and income indexing, pricing adjustments, price impact on market share, as well as the impact of economic factors on the cloud market. For each of these segments, the relations were first studied outside the model, to ensure that the functions behaved appropriately. Using placeholders and stubs, these segments were added into the model serially, to ensure that any unusual model behavior could be easily isolated. Introduction of each segment required that the model be evaluated to ensure that the reinforcing loops did not cause the model to move towards extreme outcomes. Only after assessing the model for stable behavior were we able to assess it under a variety of scenarios.

Model Validation

A number of strategies have been suggested for structural validation of system dynamics models, including boundary adequacy, structural verification, parameter verification, and dimensional consistency (Forrester & Senge 1980), (Qudrat-Ullah & Seong 2010). Boundary analysis checks if important concepts and structures are endogenous to the model. For this model, price, market share, demand, costs and revenues are considered endogenous variables. The exogenous variables include economic factors, customer demand, and coefficients for costs and revenues. Structural verification seeks to establish whether the model is consistent with available knowledge about the phenomenon under review. Since the model was assembled using constructs and causal relationships drawn from prior research in the cloud computing paradigm, this objective is met. Parameter verification is directed at establishing whether the parameters in the model are consistent with the real world phenomenon. The parameters in the cloud capacity model appear as constants and equations. Parameter values for the cost and revenue components were drawn from averages for the US. Dimensional consistency addresses whether variables and equations in the model are dimensionally consistent and correct. The model uses a mix of units, given that it addresses multiple constructs. Cloud market is measured in terms of number of customers. Processing requirements are measured in Kbps while storage requirements are modeled as GBytes. The cloud demand is measured in Kbps, while costs and revenues are measured in dollars. Behavioral analysis of the model was performed through extreme condition analysis and transition value analysis in addition to observing the model performance for chaotic and uncharacteristic behavior. No unexplained deviations were observed.

Simulation Results

Vensim® PLE, a fully functional system dynamics software package from Ventana Systems, Inc. was used to run the simulations. A medium term cloud planning horizon of 60 months was used. Longer time frames would likely involve greater uncertainty of environmental conditions and result in potentially inaccurate predictions. The experiments are conducted to validate that the model is performing realistically, and to assess the impact of different economic environments on cloud income and market share.

Scenarios Explored

The model was tested using a set of four scenarios. These represent cases of a shrinking economy, a stagnant economy, a growing economy, and a changing economy respectively and are described in Table 1. The values represent an annualized rate of change in GDP. Values for the last scenario were derived based on quarterly changes in GDP, which were then scaled up to annual values. Data was gathered from FRED (Federal Reserve Economic Research) from the Federal Reserve in St. Louis.

Scenario	Economic Factors
Shrinking Economy	−0.03
Stagnant Economy	0.0
Growing Economy	0.06
Changing Economy	[−0.08 .. 0.08] reflecting conditions during the 2008 recession

Table 1. Scenarios for Economic Conditions

All variables in the model were tracked and assessed for meaningful behavior in the context of the scenarios examined. Though evaluations were performed for each scenario, the more interesting observations are discernible when the scenarios are compared with each other. Of specific interest was how the number of cloud customers changed for different scenarios, as well as the market share for the cloud provider. Cloud income and cumulative income was also monitored. These results are illustrated in Figure 2. The graphs indicate behavior that is consistent with the conditions of the scenario as well as the decision making strategies employed in the model.

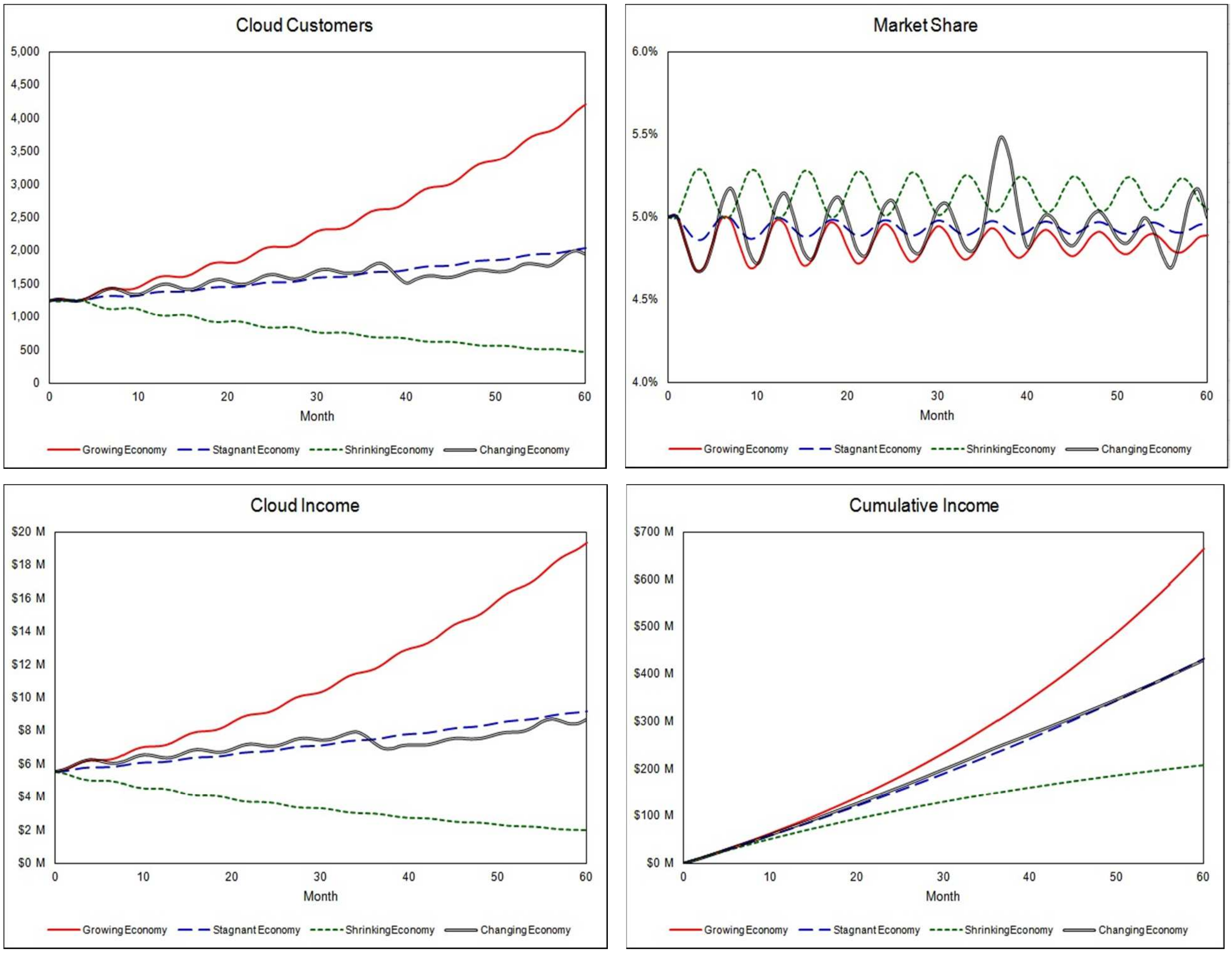

Figure 2. Simulation Results across Different Scenarios

Behavior Observed

The graph on cloud customers reflects the overall impact of economic conditions on the cloud market. Given a 6% sustained increase, it is no surprise that the overall cloud market grows significantly. This is also consistent with the projections for overall cloud market growth. For a stagnant economy, there is some growth in the cloud market, reflecting the fact that organizations offload their computing to the cloud to reduce expenses. The case for the shrinking economy leads to a substantial decline in the cloud market. The graph also indicates that the change is bumpy in all cases. This is due to the fact that there is a price correction based on income built into the model. A drop-in income triggers a price cut to boost market share. As market share and income increase, the price is increased to ensure that market share can be maintained. This adjustment accounts for the wavy pattern seen in the cloud customers.

The adjustment is better understood when observing the graph on market share. At the start of the simulation, the income is higher for the growing economy and the stagnant economy. This triggers a price increase, which slowly starts to decrease the market share. After a few periods of this behavior, the income starts to drop, and a reversal is sought through a price cut. This leads to a cyclical behavior that spans approximately 5 time periods. Also, over time, the cycle is dampened and the amplitude of the variation diminishes considerably. In the case of a shrinking economy, the reverse trend is observable. A price cut is sought immediately to grow market share. The cyclical behavior is still observed, though the damping effect is not as prominent. Also of interest is the adjustment of the price index in a manner that the market share grows over time.

Cloud income is computed for each period in the simulation. Predictably, it mirrors the cloud customers, since it determines the overall demand for services as well as the costs for the resources to provision those services. Over the period of the simulation, income rises 232% for the growing economy, 64% for the flat economy, drops 62% for the shrinking economy, and rises 52% for the changing economy. It is worth noting that though the cloud income drops in the shrinking economy scenario, its effect is not symmetric to the growing economy case, in that the drop is considerably smaller. The cumulative income reflects these findings, and smoothens the price adjustment variations. It should be borne in mind that actual cloud income would be higher if web hosting and SaaS services are also available from the cloud provider.

The results provide useful insights into the cloud computing market when a price adjusting mechanism is employed. The case of the growing and shrinking economies serve as reference points for studying growth in the cloud computing market. It is unlikely that economies will grow or shrink with that level of consistency over such an extended period. It is more likely that the changing economy scenario is more realistic, albeit with slightly better prospects than the period chosen. The simulations demonstrate healthy growth for the cloud computing industry is likely to be the case under most realistic economic conditions. This is understandable as conditions move towards an information-driven economy. Small businesses possess neither the skill set nor the budget to maintain their own data centers, and will look to cloud computing solutions to deliver customer centered applications (Marston et al. 2010). It is likely that they will be looking for SaaS solutions in addition to IaaS products, adding to cloud computing revenues.

Assessing the utility of the model can be tricky since the cloud computing phenomenon is still unfolding. The growing and shrinking economy scenarios are provided for reference, and are potentially unrepresentative in that sustained growth or shrinkage of the economy for a period of five years is unlikely. The stagnant and changing economy scenarios are more representative. When compared to actual predictions of the cloud computing demand over time (Statista 2016), the results generated by the model are fairly consistent with predictions.

Managerial and Research Implications

The cloud capacity model clearly illustrates that economic factors have major implications for the size of cloud market and attendant capacity planning. A number of insights can be gleaned through the simulations. The obvious implication is that sustained economic conditions lead to sizeable growth or contraction of the cloud market, based on the direction of the economic factors. This is hardly surprising. Nonetheless, it offers a baseline for cloud capacity planning, in that individual providers can employ appropriate tactics to ensure that they continue to be a relevant player in the cloud market.

The most vital takeaway for managers is that it is possible to ensure their continued presence in the market, under a variety of different economic conditions. The use of an income-indexed pricing policy ensures that a cloud provider can retain market share and thereby sustain income derived through the IaaS product portfolio. The effect is not immediate, but the correction occurs within a few periods, ensuring that the cloud income remains reliable and sustained, even in the face of changing economic conditions. This requires that the cloud provider adopt a dynamic pricing strategy, rather than lock in rates that have little flexibility to change. It also assumes that the cloud customers are price sensitive and will enter and leave with falling and rising cloud prices. As the market becomes more mature, cloud products will be viewed as more of a commodity with easy substitutability among products.

While the implications of the quick reactive decision are clear, the actual decision may run counter to expected thought. Thus, in the face of falling income due to economic recession, the recommendation is to cut prices in an effort to shore up market share, and thereby build income. Likewise, in times of sustained economic growth, a modest price increase stabilizes income through slightly reduced market share.

Another finding is that even in times of stagnant economic conditions, or varied economic conditions that include some contraction, cloud providers can look forward to a growing market. This is due to the fact that they provide a service that enables cost reductions for their customers. While the cloud market is not recession proof, it does enjoy a hardened set of conditions that provide it some protection in conditions where other industries are at risk. This is consistent with most cloud computing market forecasts which indicate sustained growth, even with anemic economic growth.

From a research perspective, the findings provide an array of insights about cloud capacity management. First, it demonstrates that it is possible to successfully embed dynamic decision making in the form of a reactive income-sensitive pricing policy into system dynamics models. Not only does it provide an opportunity for taking corrective action, it also provides a robust mechanism for ensuring long-term survival of the cloud provider.

Another research finding is that the reactive pricing model can be used effectively since prevailing market conditions permit it. Cloud providers have the luxury of using this option, since margins in the industry currently run above 40%, leaving them a lot of room for self-correcting actions. If margins were much tighter, this would not be an option, and tinkering with cloud prices would likely lead to loss of both market share and income.

In addition, this provides a starting point for researchers to engage in further exploration of the cloud capacity management decisions. The consideration of more diverse economic factors would form the next logical step. Additional simulations involving the speed of environment changes would shed more light on our initial findings that suggest potential linkage between cloud computing and environmental turbulence. As the cloud computing economy matures, some changes can be expected for the relationships, and growth in the market can be expected to taper off.

Conclusions

The management of cloud capacity is of critical importance for cloud providers. This research examined the relatively neglected area of cloud capacity management from the cloud provider's perspective, through the use of a system dynamics model. The model was constructed to include economic factors, cloud consumer base, cloud demand, server requirements, datacenter costs, cloud income, cloud price index, and market share. Simulations with the model indicate that cloud systems are favored in times of stagnant and varied economies. Only in the cases of sustained contracting economies are cloud providers at risk. More importantly, it demonstrates that it is possible for a cloud provider to adopt a swift reactive pricing policy that ensures long-time viability.

REFERENCES

Aceto, G., Botta, A., De Donato, W., & Pescapè, A. 2013. "Cloud monitoring: A survey," *Computer Networks* (57: 9), pp. 2093-2115.

Dillon, T., Wu, C., and Chang, E. 2010. "Cloud Computing: issues and challenges," in *proceedings of 24th IEEE International Conference on Advanced Information Networking and Applications*, pp. 27-33.

Dutta, A., Peng, G.C.A. and Choudhary, A. 2013. "Risks in enterprise cloud computing: the perspective of IT experts," *The Journal of Computer Information Systems* (53:4), pp. 39-48.

Forrester, J.W. 1961. *Industrial dynamics*, Cambridge, MA: MIT Press.

Forrester, J.W. and Senge, P.M. 1980. "Tests for building confidence in system dynamics models," *TIMS Studies in the Management Sciences* (14), pp. 209-228.

Glanz, J. 2012. "The Cloud Factories – Power, Pollution and the Internet," *New York Times*, September 22, 2012.

Hevner, A.R., March, S.T., Park, J., and Ram, S. 2004. "Design science in information systems research," *MIS Quarterly* (28:1), pp. 75-106.

Jiang, Y., Perng, C.S., Li, T., & Chang, R. 2012. "Self-adaptive cloud capacity planning," in *proceedings of the Ninth International Conference on In Services Computing (SCC 2012)*, Honolulu, Hawaii, June 24-29 , 2012, pp. 73-80.

Khansa, L. and Zobel, C.W. 2014. "Assessing Innovations in Cloud Security," *The Journal of Computer Information Systems* (54:3), pp. 45-56.

Kirk, R. 2016. "Capacity Planning for Cloud Platforms," *Century Link*, https://www.ctl.io/knowledge-base/general/capacity-in-the-cloud/, updated March 28, 2016, accessed December 1, 2016

Koomey, J.G. 2011. "Growth in Data Center Electricity Use 2005 to 2010," *Analytics Press Report*, August 2011.

Larumbe, F., and Sansò, B. 2013. "A Tabu Search Algorithm for the Location of Data Centers and Software Components in Green Cloud Computing Networks," *IEEE Transactions on Cloud Computing* (1:1), pp. 22-35.

Luftman, J., and Zadeh, H.S. 2011. "Key Information Technology and Management Issues 2010-11: An International Study," *Journal of Information Technology* (26:3), pp. 193-204.

Luna-Reyes, L.F. and Andersen, D.L. 2003. "Collecting and Analyzing Qualitative Data for System Dynamics: Methods and Models," *System Dynamics Review* (19) pp. 271–296.

Lynn, T., Van, D.W., Hunt, G. and Healy, P. 2016. "Development of a cloud trust label: a Delphi approach," *The Journal of Computer Information Systems* (56:3), pp. 185-193.

Mao, M., Li, J., & Humphrey, M. 2010. "Cloud auto-scaling with deadline and budget constraints," In *proceedings of the 11th IEEE/ACM International Conference on Grid Computing (GRID 2010)*, Brussels, Belgium, October 25-29, 2010, pp. 41-48.

Marston, S., Li, Z., Bandyopadhyay, S., and Zhang, J. 2011. "Cloud Computing – The Business Perspective," *Decision Support Systems* (51:1), pp. 176-189.

Menard, P., Gatlin, R. and Warkentin, M. 2014. "Threat protection and convenience: antecedents of cloud-based data backup," *The Journal of Computer Information Systems* (55) pp. 83-91.

Menascé, D.A., & Ngo, P. 2009. "Understanding Cloud Computing: Experimentation and Capacity Planning," In *proceedings of the 2009 Computer Measurement Group Conference*, Dallas TX, December 1-11, 2009, pp. 1-11.

Miller, R. 2009. "How Many Servers Can One Admin Manage?," *Data Center Knowledge*, December 30, 2009, http://www.datacenterknowledge.com/archives/2009/12/30/how-many-servers-can-one-admin-manage/, accessed on December 1, 2016.

Oredo, J.O., and Njihia, J. 2014. "Challenges of Cloud Computing in Business: Towards New Organizational Competencies," *International Journal of Business and Social Sciences* (5:3) pp. 150-160.

Peng, G.C.A. and Gala, C. 2014. "Cloud ERP: a new dilemma to modern organisations?," *The Journal of Computer Information Systems* (54:4), pp. 22-30.

Qudrat-Ullah, H. and Seong, B.S. 2010. "How to do structural validity of a system dynamics type simulation model: The case of an energy policy model," *Energy Policy* (38), pp. 2216-2224.

Rasmussen, N., 2015. "Calculating Space and Power Density Requirements for Data Centers, Revision 0," *American Power Conversation*, White Paper 155, Schneider Electric. March 2015.

Saripalli, P., Kiran, G. V. R., Shankar, R. R., Narware, H., & Bindal, N. 2011. "Load prediction and hot spot detection models for autonomic cloud computing," In *proceedings of the Fourth IEEE International Conference on Utility and Cloud Computing (UCC 2011)*, Melbourne, Australia, December 5-7, 2011, pp. 397-402.

Sharma, U., Shenoy, P., Sahu, S., & Shaikh, A. 2011a. "Kingfisher: Cost-aware elasticity in the cloud," In *proceeding of the 2011 IEEE Conference on Computer Communications (INFOCOM 2011)*, Shanghai, China, April 10-15, 2011, pp. 206-210.

Sharma, U., Shenoy, P., Sahu, S., & Shaikh, A. 2011b. "A cost-aware elasticity provisioning system for the cloud," In *proceedings of the 31st International Conference on In Distributed Computing Systems (ICDCS 2011)*, Minneapolis, MN, June20-14, 2011, pp. 559-570.

Statista 2016. "Total size of the public cloud computing market from 2008 to 2020," https://www.statista.com/statistics/510350/worldwide-public-cloud-computing/, accessed on October 15, 2016.

Sterman, J.D. 2000. *Business dynamics: Systems thinking and modeling for a complex world*, New York, NY: Irwin McGraw-Hill.

Wu, J., Ping, L., Ge, X., Wang, Y., & Fu, J. 2010. "Cloud storage as the infrastructure of cloud computing," In *proceedings of the 2010 International Conference on Intelligent Computing and Cognitive Informatics (ICICCI)*, Kuala Lampur, Malaysia, June 22-23, 2010, pp. 380-383.